

Bootstrapping the Mean Vector for the Observations in the Domain of Attraction of a Multivariate Stable Law

Maryam Sohrabi and Mahmoud Zarepour
University of Ottawa, Ontario, Canada

October 8, 2015

ABSTRACT

We consider a robust estimation of the mean vector for a sequence of i.i.d. observations in the domain of attraction of a stable law with different indices of stability, $DS(\alpha_1, \dots, \alpha_p)$, such that $1 < \alpha_i \leq 2$, $i = 1, \dots, p$. The suggested estimator is asymptotically Gaussian with unknown parameters. We apply an asymptotically valid bootstrap to construct a confidence region for the mean vector. A simulation study is performed to show that the estimation method is efficient for conducting inference about the mean vector for multivariate heavy-tailed distributions.

1 Introduction

Let X_1, X_2, \dots, X_n be a sequence of i.i.d. random variables from some distribution F with mean μ . Traditionally, studentization has been considered to make inference about the mean for relatively light-tailed distributions. This approach demands only the second moment assumption. Moreover, the bootstrap inference is arguably accurate and a simple approach to make inference for the univariate mean for finite variance observations (see, for example, Dickey and Efron (1996) and Singh (1981)).

Now consider that $\{X_k\}$ are in the domain of attraction of a stable law with infinite second moment. Since the sample mean \bar{X}_n is a common estimator of the mean, it is natural to base inference about μ on \bar{X}_n . Thus, there exists a constant $a_n > 0$ such that

$$S_n = a_n^{-1} \sum_{k=1}^n (X_k - \mu) \xrightarrow{d} S_\alpha,$$

where S_α is a stable random variable with index $0 < \alpha < 2$. It is known that $a_n = n^{1/\alpha} L(n)$ where L is a slowly varying functions at ∞ ; see Feller (1971) for more details. Despite the fact that the sample mean is an intuitive estimate for the population mean, the rate of convergence of the sample mean is na_n^{-1} which approaches zero very slowly when α is close to 1.

Properties of the bootstrapping for the mean of heavy-tailed distributions have been considered extensively in statistical literatures (see, for example, Hall (1990) and Knight (1989a)). It has been shown that the regular bootstrap is not consistent for estimating the distribution of the mean. For the finite variance observations, the bootstrap distribution of the sample mean converges almost surely to a fixed distribution. While, Athreya (1987) shows that the bootstrap distribution of the sample mean of infinite variance observations converges in distribution to a random probability distribution. Athreya, Lahiri, and Wu (1998) demonstrate that bootstrapping based on resampling m out of n observations, such that $m/n \rightarrow 0$, rectifies the asymptotic failure of the regular bootstrap for relatively heavy-tailed distributions. They also consider the bootstrap methods for conducting inference about the mean for a sequence of i.i.d. random variables in the domain of attraction of a stable law whose index exceeds 1. Arcones and Giné (1989) discuss almost sure and in probability bootstrap central limit theorem when the random variable X is in the domain of attraction of a stable law with infinite second moment. Hall and LePage (1996) propose a bootstrap method for estimating the distribution of the studentized mean under more general conditions on the tails of the sampling distributions. They also show that this method holds even when the sampling distribution is not in the domain of attraction of any limit law. Zarepour and Knight (1999b) consider the weak limit behavior of a certain type of point process obtained by replacing the original observations by the bootstrap sample.

We wish to make inference about the mean vector $\boldsymbol{\mu}$ of a multivariate heavy-tailed distribution. Consider the model

$$\mathbf{X}_i = \boldsymbol{\mu} + \boldsymbol{\epsilon}_i, \quad i = 1, 2, \dots, n, \quad (1.1)$$

where $\mathbf{X}_i = [X_{i1} \dots X_{ip}]^T$, $i = 1, \dots, n$, are \mathbb{R}^p -valued random vectors, and $\boldsymbol{\mu} = [\mu_1 \dots \mu_p]^T$ is an unknown fixed parameter. Let $\{\boldsymbol{\epsilon}_i\} = \{[\epsilon_{i1} \dots \epsilon_{ip}]^T\}$ form a sequences of i.i.d. random vectors with zero mean in the domain of attraction of a multivariate stable law. The following generalizes the definition of the domain of attraction of a bivariate stable law in Resnick and Greenwood (1979) to the multivariate stable law.

Definition 1. Given $\{\mathbf{X}_n = [X_{n1} X_{n2} \dots X_{np}]^T\}$ i.i.d. random vectors on \mathbb{R}^p with distribution F , let $\mathbf{S}_n = \sum_{i=1}^n \mathbf{X}_i$ and $S_n^{(j)} = \sum_{i=1}^n X_{ij}$ for $j = 1, \dots, p$. Then, $F \in DS(\alpha_1, \dots, \alpha_p)$, $\alpha_j \in (0, 2]$, if there exist sequences $\mathbf{a}_n = (a_n^{(1)}, \dots, a_n^{(p)})$, $\mathbf{b}_n \in \mathbb{R}^p$ with $a_n^{(j)} > 0$ such that

$$\left(S_n^{(1)}/a_n^{(1)}, \dots, S_n^{(p)}/a_n^{(p)} \right) - \mathbf{b}_n \xrightarrow{d} \mathbf{Y}, \quad (1.2)$$

where \mathbf{Y} is a random vector on \mathbb{R}^p with stable distribution and the distribution of \mathbf{X}_1 is in the domain of attraction of the distribution of \mathbf{Y} .

For more discussion about the class of all possible limits in (1.2) see Resnick and Greenwood (1979). For the multivariate case, observations can be in the domain of attraction of a stable law with different indices of stability. In many real life examples, some coordinates may have light tails while other coordinates may have heavier tails. In this paper, we assume that errors are in $DS(\alpha_1, \dots, \alpha_p)$ with possibly different values of $\alpha_j \in (1, 2]$ for $j = 1, \dots, p$.

It is obvious that the limiting distribution of the sample mean depends on the tail indices when the errors are in the domain of attraction of a stable law. Thus, it is hard to derive any inference for the mean vector $\boldsymbol{\mu}$ based on the limit, especially when the limiting distribution of each coordinate has different indices of stability. A bootstrap procedure may circumvent this difficulty but, as mentioned before, the ordinary bootstrap fails in this case. Using the bootstrap samples of size m , when $m/n \rightarrow 0$, typically resolves the problem. Note that the choice of m is a key issue and several investigations consider different rules to pick m . See Bickel and Sakov (2008) for more details.

These difficulties prompted us to look at a robust estimation based on M-estimate method for constructing any inference about the mean vector such as confidence regions when the errors are in the domain of attraction of a multivariate stable law with possibly different indices of stability in $(1, 2]$. In our approach, the proposed robust estimation method of the mean vector has higher rate of convergency compared to the sample mean. We also show that the regular bootstrap is applicable since the limiting distribution is a multivariate normal distribution.

Section 2 presents our main theorem, the robust estimation of the mean vector for a sequence of i.i.d. observations in the domain of attraction of a stable law with different indices of stability, $DS(\alpha_1, \dots, \alpha_p)$, such that $1 < \alpha_i \leq 2$, $i = 1, \dots, p$. The bootstrap procedure is discussed in Section 3. Section 4 presents some simulations supporting the results of this paper.

2 M-estimates of the mean vector

Let $\boldsymbol{\mu}$ be the parameter vector of interest and for a random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$, let \mathbf{X} denote the data satisfying (1.1). The classical M-estimate for $\boldsymbol{\mu}$, denoted by $\hat{\boldsymbol{\mu}}_M$, is defined as the minimizer of the function

$$\sum_{i=1}^n (\rho(\mathbf{X}_i - \boldsymbol{\beta}) - \rho(\boldsymbol{\epsilon}_i)),$$

with respect to $\boldsymbol{\beta} = [\beta_1 \dots \beta_p]^T$, where ρ is an almost everywhere differentiable convex function. This guarantees the uniqueness of the solution. For more details see Davis, Knight, and Liu (1992). For convenience, similar to Zarepour and Roknossadati (2008), we consider the multivariate loss function as

$$\rho(x_1, \dots, x_p) = \rho_1(x_1) + \dots + \rho_p(x_p), \quad (2.1)$$

where ρ_j , $j = 1, \dots, p$, are univariate loss functions. A good justification for using the objective function of the form (2.1) is the ability to calibrate each coordinate to derive more precise estimates in practice.

Here, we impose the following assumptions on the functions ρ_j , for $j = 1, \dots, p$.

- A1. $\rho_j : \mathbb{R} \rightarrow \mathbb{R}$ is a convex and twice differentiable function, and take $\psi_j = \rho'_j$, and $\psi'_j = \rho''_j$.
- A2. $E(\psi_j) = 0$, $E(\psi_j^2) < \infty$, and $0 < |E(\psi'_j)| < \infty$.

A3. ψ_j has Lipschitz-continuous derivative ψ'_j ; i.e., there exists a real constant $k \geq 0$ such that for all x and y ,

$$|\psi'_j(x) - \psi'_j(y)| \leq k|x - y|.$$

The following lemma is used to prove our main results.

Lemma 1. *Suppose that $\{S_n(\cdot)\}$ is a sequence of convex stochastic processes on \mathbb{R} and suppose that*

$$S_n(\cdot) \xrightarrow{d} S(\cdot).$$

Then $\{S_n(\cdot)\}$ has a unique minimum κ_n . If κ minimize $S(\cdot)$, then

$$\kappa_n \xrightarrow{d} \kappa.$$

Proof. The proof is given in Lemma 2.2 of Davis et al. (1992). See also page 279 from Knight (1989b). □

Theorem 1. *Suppose (1.1) holds. With the loss function (2.1), let $\hat{\mu}_M$ be the M -estimation of the mean vector for a sequence of i.i.d. observations in the domain of attraction of a stable law with indices of stability $(\alpha_1, \dots, \alpha_p)$ such that $1 < \alpha_j \leq 2$, $j = 1, \dots, p$. Then, we have*

$$\mathbf{W}_n = \sqrt{n}(\hat{\mu}_M - \mu) \xrightarrow{d} \mathbf{W}, \quad (2.2)$$

where $\mathbf{W} = [W_1 \dots W_p]^T$ has a multivariate normal distribution with mean zero and covariance matrix $\Sigma = \text{diag} \left(\frac{E[(\psi_1(\epsilon_{11}))^2]}{E^2(\psi'_1(\epsilon_{11}))}, \dots, \frac{E[(\psi_p(\epsilon_{1p}))^2]}{E^2(\psi'_p(\epsilon_{1p}))} \right)$.

Proof: Under the conditions A1-A3, define the convex process

$$\begin{aligned} A_n(u_1, \dots, u_p) &= \sum_{j=1}^p \sum_{i=1}^n \left(\rho_j \left(\epsilon_{ij} - n^{-1/2} u_j \right) - \rho_j(\epsilon_{ij}) \right) \\ &= \frac{-1}{\sqrt{n}} \sum_{j=1}^p u_j \sum_{i=1}^n \psi_j(\epsilon_{ij}) + \frac{1}{2n} \sum_{j=1}^p u_j^2 \sum_{i=1}^n \psi'_j(c_{ij}), \end{aligned} \quad (2.3)$$

where $u_j = n^{1/2}(\hat{\mu}_{Mj} - \mu_j)$, for $j = 1, \dots, p$, and c_{ij} is between ϵ_{ij} and $\epsilon_{ij} - n^{-1/2} u_j$. Asymptotically, $\psi'_j(c_{ij})$ can be replaced by $\psi'_j(\epsilon_{ij})$ in (2.3) since

$$n^{-1} \sum_{j=1}^p |\psi'_j(\epsilon_{ij}) - \psi'_j(c_{ij})| \leq kn^{-1} \sum_{j=1}^p |n^{-1/2} u_j| \xrightarrow{P} 0.$$

It is well known that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_j(\epsilon_{ij}) \xrightarrow{d} E[(\psi_j(\epsilon_{1j}))^2]^{1/2} Z_j,$$

where $Z_j, j = 1, \dots, p$, have standard normal distributions. Therefore,

$$A_n(u_1, \dots, u_p) \xrightarrow{d} A(u_1, \dots, u_p) = - \sum_{j=1}^p u_j E[(\psi_j(\epsilon_{1j}))^2]^{1/2} Z_j + \frac{1}{2} \sum_{j=1}^p u_j^2 E(\psi_j'(\epsilon_{1j})).$$

From Lemma 1, the minimizer of A_n converges to the minimizer of A . Thus, (2.2) follows by setting the derivative of $A(u_1, \dots, u_p)$ to 0 and solving for u_1, u_2, \dots , and u_p . Note that $\mathbf{W} = [W_1 \dots W_p]^T$ in (2.2) has a multivariate normal distribution and

$$W_j = \frac{E[(\psi_j(\epsilon_{1j}))^2]^{1/2}}{E(\psi_j'(\epsilon_{1j}))} Z_j,$$

$j = 1, \dots, p$, are independent. □

Based on Theorem 1, a simple approach to construct a $100(1 - \alpha)\%$ confidence region for the mean of a p -dimensional random vectors in the domain of attraction of a multivariate stable law with large sample size is the ellipsoid determined by all $\boldsymbol{\mu}$ such that

$$CR_{1-\alpha} = \{\boldsymbol{\mu} : n(\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu})^T \mathbf{S}^{-1}(\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu}) \leq \tau_{1-\alpha}\}. \quad (2.4)$$

Here, $\tau_{1-\alpha}$ is the $1 - \alpha$ quantile of a $\chi^2(p)$ distribution and \mathbf{S} is the estimated value of Σ using residuals, $\mathbf{e}_i = \mathbf{X}_i - \hat{\boldsymbol{\mu}}_M$ for $i = 1, 2, \dots, n$.

The confidence region in (2.4) gives the joint knowledge concerning reasonable values of $\boldsymbol{\mu}$ when the correlation between the measured variables is taken into account. Typically, any summary of conclusions includes confidence statements about the individual component means specially when the covariance matrix is diagonal similar to our case. Let (1.1) hold for i.i.d. random vectors $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$. Consider the following linear combination

$$\mathbf{a}^T \mathbf{X} = a_1 \mathbf{X}_1 + a_2 \mathbf{X}_2 + \dots + a_p \mathbf{X}_p.$$

Simultaneously for all \mathbf{a} , the interval

$$\left(\mathbf{a}^T \hat{\boldsymbol{\mu}}_M - \sqrt{\tau_{1-\alpha} \mathbf{a}^T \mathbf{S} \mathbf{a}}, \mathbf{a}^T \hat{\boldsymbol{\mu}}_M + \sqrt{\tau_{1-\alpha} \mathbf{a}^T \mathbf{S} \mathbf{a}} \right)$$

contains $\mathbf{a}^T \boldsymbol{\mu}$ with probability $1 - \alpha$. The consecutive choices $\mathbf{a}^T = (1, 0, \dots, 0)$, $\mathbf{a}^T = (0, 1, \dots, 0)$, and so on through $\mathbf{a}^T = (0, 0, \dots, 1)$ for the χ^2 -intervals allow us to conclude that

$$\begin{aligned} \hat{\mu}_{M1} - \sqrt{\tau_{1-\alpha} s_{11}} &\leq \mu_1 \leq \hat{\mu}_{M1} + \sqrt{\tau_{1-\alpha} s_{11}} \\ \hat{\mu}_{M2} - \sqrt{\tau_{1-\alpha} s_{22}} &\leq \mu_2 \leq \hat{\mu}_{M2} + \sqrt{\tau_{1-\alpha} s_{22}} \\ &\vdots \\ \hat{\mu}_{Mp} - \sqrt{\tau_{1-\alpha} s_{pp}} &\leq \mu_p \leq \hat{\mu}_{Mp} + \sqrt{\tau_{1-\alpha} s_{pp}} \end{aligned}$$

all hold simultaneously with confidence coefficient $1 - \alpha$.

3 Bootstrapping the Mean Vector

It has been pointed out that the regular bootstrap fails to estimate the distribution for the mean of heavy-tailed observations. The main reason for the failure of the regular bootstrap comes from the fact that the rare events occur when we resample the data. This means the resampling procedure will remember the magnitude of the observations in the resampled data. This fact can be reflected in the point process theory which is used as a machinery tool for the asymptotic theory of heavy-tailed observations. Let $\mathbf{X}_1, \mathbf{X}_2, \dots$ be a sequence of i.i.d. random vectors in $DS(\alpha_1, \dots, \alpha_p)$. Given $\mathbf{X}_1, \dots, \mathbf{X}_n$, we draw an i.i.d. sequence of observations $\mathbf{X}_1^*, \dots, \mathbf{X}_n^*$ from the empirical distribution

$$F_n(\cdot) = \frac{1}{n} \sum_{i=1}^n \varepsilon_{\mathbf{X}_i}(\cdot).$$

Define $M_{i,n}^* = \sum_{k=1}^n I(\mathbf{X}_k^* = \mathbf{X}_i)$. By Lemma 3.1 of Zarepour and Knight (1989a), we have

$$(M_{1,n}^*, \dots, M_{n,n}^*, 0, 0, \dots) \xrightarrow{d} (M_1^*, \dots, M_n^*, \dots),$$

where M_1^*, M_2^*, \dots are i.i.d. Poisson(1) random variables. Define the point process

$$\xi_n = \sum_{k=1}^n \varepsilon_{(k/n, a_n^{-1} \epsilon_k)},$$

where ε_x is a measure defined by $\varepsilon_x(A) = I(x \in A)$ for any Borel set A such that $A \subseteq [0, 1] \times \mathbb{R}$. Also, let $\mathbf{X}_n^* = [X_{i1}^* \dots X_{ip}^*]^T$. Thus, we have

$$\sum_{i=1}^n \varepsilon_{((a_n^{(1)})^{-1} X_{i1}^*, \dots, (a_n^{(p)})^{-1} X_{ip}^*)} = \sum_{i=1}^n M_{i,n}^* \varepsilon_{((a_n^{(1)})^{-1} X_{i1}, \dots, (a_n^{(p)})^{-1} X_{ip})}.$$

With a considerable help from Theorem 4 of Resnick and Greenwood (1979) and Resnick (2004) and Zarepour and Knight (1999b), it can be shown that

$$\sum_{i=1}^n \varepsilon_{((a_n^{(1)})^{-1} X_{i1}^*, \dots, (a_n^{(p)})^{-1} X_{ip}^*)} \xrightarrow{d} \sum_{i=1}^{\infty} M_i^* \varepsilon_{(\text{sign}(\gamma_{i1})|\gamma_{i1}|^{1/\alpha_1} \Gamma_i^{-1/\alpha_1}, \dots, \text{sign}(\gamma_{ip})|\gamma_{ip}|^{1/\alpha_p} \Gamma_i^{-1/\alpha_p})}, \quad (3.1)$$

in distribution. Here, $\{\Gamma_1, \Gamma_2, \dots\}$ is a sequence of arrival times of a Poisson process with unit arrival rate and $\gamma_i = (\gamma_{i1}, \dots, \gamma_{ip}) \sim G$ and G is a distribution on the boundary of unit sphere. To have a valid bootstrap, we expect to have $M_i^* = 1$ which is not the case here.

The limiting distribution for the bootstrap sample mean, $\bar{\mathbf{X}}^*$, can be derived from (3.1) and continuous mapping theorem along with some other mathematical discussions. Similar to the univariate case, this result shows that the regular bootstrap fails asymptotically. As discussed in the introduction, a subsampling scheme (m out of n bootstrap such that $m/n \rightarrow 0$) is an appropriate approach to achieve asymptotic validity of a bootstrap procedure for constructing a confidence region for the

mean vector of i.i.d. heavy-tailed data. However, choosing the proper subsample size m is of great concern to many authors.

On the other hand, Theorem 1 shows that the weak limit behavior of $\sqrt{n}(\hat{\boldsymbol{\mu}}_M - \boldsymbol{\mu})$ is a multivariate normal distribution. Thus, the regular bootstrap works if we use the robust estimates (M-estimates) for the mean vector. Our approach in this section is to consider a bootstrap approach to estimate the confidence region for $\boldsymbol{\mu}$. Given $\mathbf{X} = [\mathbf{X}_1 \dots \mathbf{X}_n]$, find M-estimates of $\boldsymbol{\mu}$ in model (1.1) using the objective function in (2.1). Then calculate the residuals, where

$$\mathbf{e}_i = \mathbf{X}_i - \hat{\boldsymbol{\mu}}, \quad i = 1, 2, \dots, n. \quad (3.2)$$

Let $\mathbf{e}_1^*, \dots, \mathbf{e}_n^*$ be a sample of size n from the centered residuals in (3.2). These assumptions imply the following lemma.

Lemma 2. *Let $\{\mathbf{e}_1^*, \dots, \mathbf{e}_n^*\}$ be an i.i.d. sample from $F_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n I(\mathbf{e}_i - \bar{\mathbf{e}} \leq \mathbf{x})$ where $\mathbf{e}_i^* = [e_{i1}^* \dots e_{ip}^*]^T$, $i = 1, \dots, n$, and E^* denotes the expectation under F_n . Also, let ψ_j , $j = 1, \dots, p$, be an odd function and satisfy conditions A2-A3. Then, for $j = 1, \dots, p$, we have*

$$(i) \quad E^*(\psi_j(\mathbf{e}_{1j}^*)) = 0.$$

$$(ii) \quad n^{-1/2} \hat{\sigma}_j^{-1} \sum_{i=1}^n \psi_j(\mathbf{e}_{ij}^*) \xrightarrow{d} Z_j, \text{ in probability, where } Z_j \text{ has a standard normal distribution} \\ \text{and } \hat{\sigma}_j^2 = E^*(\psi_j^2(\mathbf{e}_{1j}^*)) = \frac{1}{n} \sum_{i=1}^n \psi_j^2(\mathbf{e}_{ij}^*) \xrightarrow{p} E(\psi_j^2(\epsilon_{1j})).$$

$$(iii) \quad E^*(\psi_j'(\mathbf{e}_{1j}^*)) = \frac{1}{n} \sum_{i=1}^n \psi_j'(\mathbf{e}_{ij}^*) \xrightarrow{p} E(\psi_j'(\epsilon_{1j})).$$

Proof: The proofs are straightforward under the conditions that ψ_j is odd and the bootstrap errors are symmetric. We omit the proofs here; for more details see Singh (1981). □

Now, we find $\{\mathbf{X}_i^*\}_{i=1}^n$ from the model $\mathbf{X}_i^* = \hat{\boldsymbol{\mu}} + \mathbf{e}_i^*$. Then, we have

$$\begin{aligned} A_n^*(u_1^*, \dots, u_p^*) &= \sum_{i=1}^n (\rho(\mathbf{X}_i^* - \hat{\boldsymbol{\mu}}) - \rho(\mathbf{e}_i^*)) \\ &= \frac{-1}{\sqrt{n}} \sum_{j=1}^p u_j^* \sum_{i=1}^n \psi_j(\mathbf{e}_{ij}^*) + \frac{1}{2n} \sum_{j=1}^p u_j^{*2} \sum_{i=1}^n \psi_j'(c_{ij}^*), \end{aligned}$$

where $u_j^* = n^{1/2}(\hat{\mu}_j^* - \hat{\mu}_j)$ for $j = 1, \dots, p$ and c_{ij}^* is between \mathbf{e}_{ij}^* and $\mathbf{e}_{ij}^* - n^{-1/2}u_j^*$. Then by Lemma 2, we have

$$\mathbf{W}^* = \sqrt{n}(\hat{\boldsymbol{\mu}}^* - \hat{\boldsymbol{\mu}}) \xrightarrow{d} \mathbf{W}, \quad (3.3)$$

in probability, where \mathbf{W} is defined in (2.2). We apply (3.3) to approximate the critical points of \mathbf{W} . To do so, we carry out a large number, say B , of the bootstrap replicates of size n from

$$C^* = n(\hat{\boldsymbol{\mu}}_M^* - \hat{\boldsymbol{\mu}})^T S^{*-1} (\hat{\boldsymbol{\mu}}_M^* - \hat{\boldsymbol{\mu}}).$$

Set $\hat{\tau}_\alpha$ to be the 100α -th percentile value of $\{C^*(b), b = 1, 2, \dots, B\}$. Thus an approximate confidence region for μ at level $100(1 - \alpha)\%$ will be

$$n(\hat{\mu}_M - \mu)^T \mathbf{S}^{-1}(\hat{\mu}_M - \mu) \leq \hat{\tau}_{1-\alpha}. \quad (3.4)$$

4 Simulation

To illustrate the preceding results, some simulation studies are performed. The first step in our simulation study is choosing the loss function. An example for the univariate $\rho_j(\cdot)$ is the Huber loss function given by

$$\rho_j(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leq c_j, \\ c_j|x| - \frac{1}{2}c_j^2 & \text{if } |x| > c_j, \end{cases} \quad (4.1)$$

for a known constant c_j ; see Huber (1981). Then, $\psi_j(x) = \max[\min(x, c_j), -c_j]$. The Huber loss function satisfies conditions A1-A3 except ψ_j might not exist everywhere. In this case, although ψ_j is not differentiable at a countable number of points, the results will usually hold with some additional complexity in the proofs.

The choice of a truncation value c_j is of practical interest specially when we have different indices of stability. The following univariate simulation study is undertaken in order to explore whether there is a relationship between values of c_j in the Huber loss function and the index of stability α .

Consider the univariate model

$$X_i = \mu + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (4.2)$$

where $\{\epsilon_i\} \in DS(1 < \alpha \leq 2)$. We generate the time series $\{X_i\}_{i=1}^n$ in model (4.2) for $\mu = 3$ and $n = 100$ with different values of $1 < \alpha \leq 2$. Then, the M-estimates of μ in (4.2) are calculated from generated time series using the Huber loss function given in (4.1). To seek a more efficient c in the Huber loss function, the estimation is repeated for different values of c between 0.5 and 4.5 for each choice of α . We find the average deviation by calculating the absolute deviation $|\hat{\mu}_M - \mu|$ and then carrying out 10,000 replications. The numbers in Table 1 are the averages of the replications. Meanwhile, the scatterplot of the average deviations presented in Figure 1. For each level of α and c , the minimum of the average deviations are emboldened in Table 1. This table shows that, for instance, for $1.1 \leq \alpha \leq 1.4$ if we choose $c = 1$, we get the minimum of the error estimation. Table 1 and Figure 1 also show that there is a positive relationship between the truncation value c and the index of stability α . In fact, to have less estimation errors, we must choose larger value of c as α gets larger.

Now consider the bivariate model

$$\begin{bmatrix} X_{i1} \\ X_{i2} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{i1} \\ \epsilon_{i2} \end{bmatrix}, \quad i = 1, 2, \dots, n. \quad (4.3)$$

Set $[\mu_1 \ \mu_2]^T = [1 \ 14]^T$ and $\{\epsilon_i\}$ are in a domain of attraction of a symmetric bivariate stable laws with indices of stabilities $1 < \alpha_1 \leq 2$ and $1 < \alpha_2 \leq 2$. To generate $\{\epsilon_i\} = \{[\epsilon_{i1} \ \epsilon_{i2}]^T\}$

Table 1: Average values of $|\hat{\mu}_M - \mu|$ for different values of α and truncation values of c in the Huber loss function with the replication size of 10,000.

α	c								
	0.5	1	1.5	2	2.5	3	3.5	4	4.5
1.1	0.123	0.126	0.134	0.143	0.153	0.162	0.171	0.180	0.188
1.2	0.125	0.125	0.131	0.138	0.146	0.154	0.161	0.169	0.176
1.3	0.128	0.126	0.127	0.136	0.142	0.149	0.156	0.162	0.168
1.4	0.128	0.125	0.127	0.131	0.137	0.142	0.147	0.153	0.158
1.5	0.129	0.125	0.125	0.128	0.132	0.136	0.140	0.145	0.148
1.6	0.130	0.128	0.123	0.125	0.126	0.131	0.134	0.137	0.141
1.7	0.130	0.124	0.122	0.122	0.123	0.125	0.127	0.130	0.132
1.8	0.129	0.122	0.120	0.118	0.118	0.119	0.121	0.122	0.124
1.9	0.131	0.124	0.120	0.118	0.117	0.118	0.118	0.119	0.120
2.0	0.131	0.123	0.119	0.116	0.115	0.114	0.114	0.114	0.114

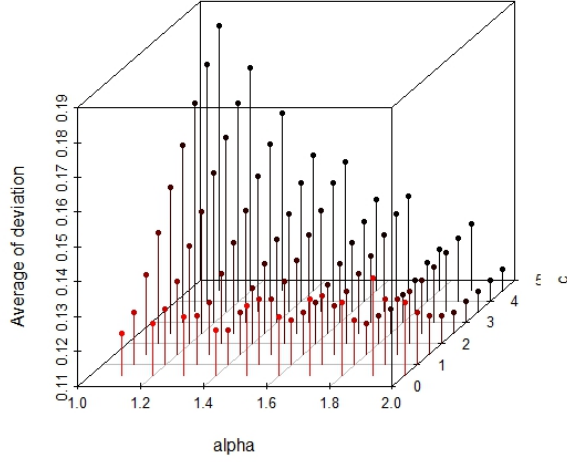


Figure 1: Average values of $|\hat{\mu}_M - \mu|$ for different values of α and truncation values of c in the Huber loss function with the replication size of 1,000.

in (4.3) with the preceding indices of stabilities, consider the set of $K = 10,000$ points $\{\gamma_i = (\cos \theta_i, \sin \theta_i) : \theta_i \in [0, 2\pi], i = 1, \dots, 10000\}$ on the boundary of the unit circle. By (3.1), we

draw the error $\{\epsilon_1\}$ from

$$\epsilon_1 = [\epsilon_{11} \ \epsilon_{12}]^T = \left[\sum_{i=1}^K \text{sign}(\gamma_{i1}) |\gamma_{i1}|^{1/\alpha_1} \Gamma_i^{-1/\alpha_1}, \sum_{i=1}^K \text{sign}(\gamma_{i2}) |\gamma_{i2}|^{1/\alpha_2} \Gamma_i^{-1/\alpha_2} \right]^T, \quad (4.4)$$

where $\Gamma_i = E_1 + \dots + E_i$ and $E_j, j \geq 1$ are i.i.d. random variables with a standard exponential distribution. To get the exact value of the innovations, K must tend to ∞ . Perform this procedure again n times independently to generate random numbers $\{\epsilon_i\}, i = 1, \dots, n$.

To acquire an intuitive feeling of the bivariate observations with different indices of stability, we generate errors from (4.4) with indices of stability $(\alpha_1, \alpha_2) = (1.3, 1.8)$ and $(2.0, 1.2)$ and sample size $n = 1,000$. The observations $[X_{i1} \ X_{i2}]^T, i = 0, 1, \dots, n$, are simulated from (4.3), and based on these observations, we plot the joint density of $[X_{i1} \ X_{i2}]^T, i = 1, \dots, n$. Figure 2 and 3 present the joint density of $[X_{i1} \ X_{i2}]^T$ when $(\alpha_1, \alpha_2) = (1.3, 1.8)$ and $(\alpha_1, \alpha_2) = (2.0, 1.2)$, respectively.

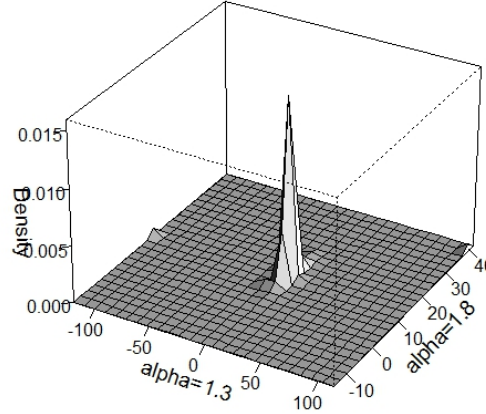


Figure 2: Density plot for the bivariate observations $[X_{i1} \ X_{i2}]^T$ given in (4.3) where $\alpha_1 = 1.3$ and $\alpha_2 = 1.8$.

To illustrate the results of Theorem 1, we perform the following simulation study to conduct $100(1 - \alpha)\%$ confidence region for the mean vector in model (4.3). All the corresponding distributions of the innovations come from symmetric bivariate stable laws with indices of stability $(\alpha_1, \alpha_2) = (1.2, 1.1), (1.5, 1.5), (1.5, 1.9), (1.3, 1.8)$, and $(2.0, 1.2)$ with sample sizes $n = 100, 200$, and 500 . The simulation scheme for each choice of n and (α_1, α_2) is as follows:

- (i) Generate $\{\epsilon_i\}$ in (4.3) with the preceding indices of stabilities.
- (ii) Find $\{\mathbf{X}_i\}_{i=1}^n$ from (4.3). Then estimate $\boldsymbol{\mu}$ by $\hat{\boldsymbol{\mu}}_M$ using the bivariate convex function ρ given in (2.1) and apply the Huber loss function in (4.1) for ρ_1 and ρ_2 . Note that, according to the

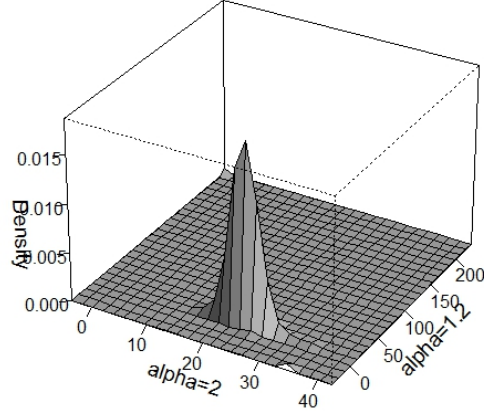


Figure 3: Density plot for the bivariate observations $[X_{i1} \ X_{i2}]^T$ given in (4.3) where $\alpha_1 = 2$ and $\alpha_2 = 1.2$.

values of α_1 and α_2 , the values of c_1 and c_2 are chosen from Table 2.1 and plugging them in the result in Theorem 1.

- (iii) Estimate Σ by using the residuals, $e_i = \mathbf{X}_i - \hat{\boldsymbol{\mu}}_M$, $i = 1, 2, \dots, n$.
- (iv) Estimate the naive $(1 - \alpha)$ -percentiles from 3,000 bootstrap replications of $C^* = n(\hat{\boldsymbol{\mu}}_M^* - \hat{\boldsymbol{\mu}})^T S^{*-1} (\hat{\boldsymbol{\mu}}_M^* - \hat{\boldsymbol{\mu}})$. To do so, draw a sample of size n from centered residuals denoted by $\hat{\epsilon}_1^*, \dots, \hat{\epsilon}_n^*$, and find $\{X_i^*\}_{i=1}^n$ from (4.3). Then estimate the parameters $[\mu_1 \ \mu_2]^T$ using the bootstrap observations by the same minimization of the objective function used in step (ii). Out of 3,000 bootstrap replications, compute the $(1 - \alpha)$ quantile of $C^*(b)$, $b = 1, 2, \dots, 3000$.
- (v) Compute the confidence region from (3.4).

This experiment was repeated 8,000 times to estimate the coverage probabilities by checking whether $[\mu_1 \ \mu_2]^T = [1 \ 14]^T$ are located within the estimated confidence intervals or not. Table 2 presents the coverage probabilities for different combinations of α_1 and α_2 when we set confidence levels equal to 0.90, 0.95 and 0.99. As seen in Table 2, the M-estimates provide significantly precise estimation such that the coverage probabilities for their related confidence intervals are fairly close to the confidence levels.

Table 2: Estimated coverage probabilities by employing different choices of sample sizes, indices of stability, and confidence levels

(α_1, α_2)	$n = 100$			$n = 200$			$n = 500$		
	$I_{0.90}$	$I_{0.95}$	$I_{0.99}$	$I_{0.90}$	$I_{0.95}$	$I_{0.99}$	$I_{0.90}$	$I_{0.95}$	$I_{0.99}$
(1.2,1.1)	0.886	0.944	0.987	0.889	0.948	0.987	0.898	0.948	0.988
(1.5,1.5)	0.890	0.945	0.986	0.903	0.948	0.989	0.902	0.946	0.986
(1.5,1.9)	0.888	0.939	0.983	0.900	0.947	0.991	0.891	0.947	0.990
(1.3,1.8)	0.885	0.943	0.984	0.900	0.950	0.988	0.889	0.949	0.986
(2.0,1.2)	0.892	0.942	0.987	0.893	0.945	0.988	0.900	0.943	0.990

5 Some notes and remarks

In this paper, a robust estimation of the mean vector is considered when the observations are a sequence of i.i.d. random vectors in the domain of attraction of a stable law with different indices of stability, $DS(\alpha_1, \dots, \alpha_p)$, such that $1 < \alpha_i \leq 2$, $i = 1, \dots, p$.

Notice that it is not necessary to enforce the condition that $1 < \alpha \leq 2$. For $0 < \alpha \leq 1$, $E(\mathbf{X}_i)$, $i = 1, \dots, n$, does not exist. Therefore, $\boldsymbol{\mu}$ is not the mean for the observations $\{\mathbf{X}_i\}$ in (1.1) but we can still consider $\boldsymbol{\mu}$ as the location parameter. The estimation procedure and the asymptotic behaviour of our M-estimate remains valid and the bootstrapping still works when we wish to estimate the location parameter $\boldsymbol{\mu}$. In this case $\{\epsilon_i\}$, $i = 1, \dots, n$ in model (1.1) should be considered to be in $DS(\alpha_1, \dots, \alpha_p)$ with $\epsilon_{1j} \stackrel{d}{=} -\epsilon_{1j}$ for $j = 1, \dots, p$.

The model (1.1) can be modified to a more complex form when $\boldsymbol{\mu}$ plays the role of the location parameter in other statistical methods. As an example, one may consider the time series models with the location parameter, $\mathbf{X}_t = \boldsymbol{\mu} + \Phi(B)\mathbf{X}_{t-1} + \epsilon_t$. Thus, such analysis can help to characterize the asymptotic distribution for the location parameter in more complex statistical models. Another application for the location parameter is calculating the data depth of the observations in the domain of attraction of a multivariate stable law (see Liu, Parelius, and Singh (1999)).

REFERENCES

- [1] Arcones, M. A. and Giné, E. (1989) The bootstrap of the mean with arbitrary sample size. *Annales de l'Institut Henri Poincaré*, 25, 457-481.
- [2] Athreya, K.B. (1987) Bootstrap of the mean in the infinite variance case. *Annals of Statistics*, 15, 724-731.
- [3] Athreya, K.B., Lahiri, S.N., and Wu, W. (1998) Inference for heavy tailed distributions. *Journal of Statistical Planning and Inference*, 66, 61-75.
- [4] Bickel, P.J. and Sakov, A. (2008) On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica*, 18, 967-985.

- [5] Diccio, T.T. and Efron, B. (1996) Bootstrap confidence intervals. *Statistical Science*, 11, 180-228.
- [6] Davis, R.A., Knight, K., and Liu, J. (1992) M-estimation for autoregressions with infinite variance. *Stochastic Processes and Their Applications*, 40, 145-180.
- [7] Feller, W. (1971) *An Introduction to Probability Theory and its Applications*. Volume 2, second edition, Wiley, New York.
- [8] Hall, P. (1990) Asymptotic properties of the bootstrap for heavy-tailed distributions. *Annals of Probability*, 18, 1342-1360.
- [9] Haan, L. de and Resnick, S. I. (1977) Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheor. Verw. Gebiete*, 40, 317-337.
- [10] Hall, P. and LePage, R. (1996) On bootstrap estimation of the distribution of the studentized mean. *Annals of the Institute of Statistical Mathematics*, 48, 403-421.
- [11] Huber, P.J. (1981) *Robust statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.
- [12] Knight, K. (1989a). On the bootstrap of the sample mean in the infinite variance case. *Annals of Statistics*. 17, 1168-1175.
- [13] Knight, K. (1989b) Limit theory for autoregressive-parameter estimates in an infinite-variance random walk. *Canadian Journal of Statistics*, 17, 261-278.
- [14] Liu, R.Y., Parelius, J.M., and Singh, K (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics*, 27, 783-858.
- [15] Resnick, S. (2004) On the foundations of multivariate heavy-tail analysis. *Journal of Applied Probability*, 41, 191-212.
- [16] Resnick, S. and Greenwood, P. (1979) A bivariate stable characterization and domains of attraction. *Journal of Multivariate Analysis*, 9, 206-221.
- [17] Singh, K. (1981) On the asymptotic accuracy of efron bootstrap. *The Annals of Statistics*, 9, 1187-1195.
- [18] Zarepour, M. and Knight, K. (1999a) Bootstrapping unstable first order autoregressive process with errors in the domain of attraction of stable law. *Stochastic Models*, 15, 11-27.
- [19] Zarepour, M. and Knight, K. (1999b) Bootstrapping point processes with some applications. *Stochastic processes and their applications*, 84, 81-90.
- [20] Zarepour, M. and Roknossadati, S.M. (2008) Multivariate autoregression of order one with infinite variance innovations. *Econometric Theory*, 24, 677-695.